Computing the equivalent number of parameters of fixed-interval smoothers

Giancarlo Ferrari-Trecate\textsuperscript{a,1} and Giuseppe De Nicolao\textsuperscript{a}

\textsuperscript{a}Dipartimento di Informatica e Sistemistica, Università degli Studi di Pavia, Via Ferrata 1, 27100 Pavia, Italy
Phone: +39.382.505484, Fax: +39.382.505373
E-mail: \texttt{\{denicolao,ferrari\}@conpro.unipv.it}

\textbf{Abstract}

The problem of reconstructing an unknown signal from \( n \) noisy samples can be addressed by means of nonparametric estimation techniques such as Tikhonov regularization, Bayesian regression and state-space fixed-interval smoothing. The practical use of these approaches calls for the tuning of a regularization parameter that controls the amount of smoothing they introduce. The leading tuning criteria, including Generalized Cross Validation and Maximum Likelihood, involve the repeated computation of the so-called equivalent number of parameters \( q(\gamma) \), a normalized measure of the flexibility of the nonparametric estimator. The paper develops new state-space formulas for the computation of \( q(\gamma) \) in \( O(n) \) operations. The results are specialized to the case of uniform sampling yielding closed-form expressions of \( q(\gamma) \) for both linear splines and first-order deconvolution.

\textbf{Key words}: Inverse problems, neural networks, regularization, Bayesian estimation, deconvolution, Kalman filter.

\section{Introduction}

In engineering and science one is often faced with the problem of reconstructing off-line a continuous function of a real variable (e.g. time) given a finite number of noisy samples. Notable examples include the reconstruction of nonlinear maps from discrete measurements and also inverse problems such as the deconvolution problems arising in model identification (Magni et al., 2001). Corresponding author.
al. 1998), physics (Bertero 1989), physiology (De Nicolao et al. 1997), DNA segmentation (Ives et al. 1994), etc.. In this paper, the attention is restricted to those problems that involve the use of linear finite-dimensional operators either to model the unknown signal or describe the physical system to whom it is applied. Rather interestingly, there are three major nonparametric approaches to the problem that, though under different names, are substantially equivalent.

The first one, known as Tikhonov regularization (Tikhonov and Arsenin 1977), adopts a deterministic viewpoint and finds an estimate through the minimization of a cost functional which is the sum of the usual sum-of-squares fit and a “regularization term” penalizing the irregularity of the solution. For instance the integral of the squared first or second derivative of the solution may be used, in which case linear or cubic smoothing splines are obtained if the signal to be reconstructed is directly measured (Wahba 1990). The trade off between data-fit and smoothness of the solution depends on the relative importance of the two terms in the cost functional and is controlled by a positive scalar $\gamma$ called ”regularization parameter”. It is widely recognized that a careful tuning of $\gamma$ is essential in order to obtain a satisfactory estimate and several criteria have been proposed, the most popular of which is Generalized Cross Validation (GCV) (Golub et al. 1979).

The second approach, namely Bayesian regression (MacKay 1992), (Williams and Rasmussen 1996)), adopts a probabilistic viewpoint and describes the unknown signal as a random process whose probability distribution reflects the available prior knowledge. Then, the solution is just the posterior expectation of the signal given the noisy observations. Under Gaussianity assumptions, the user has only to specify the noise statistics and the autocovariance function of the unknown signal. The estimate is obtained through the solution of a system of $n$ linear equations, whose computational complexity is $O(n^3)$. Bayesian regression is equivalent to Tikhonov regularization in the sense that, given any regularized solution, there exist statistical priors such that the same solution is obtained through Bayesian regression (Wahba 1990). It would seem that Bayesian regression is more direct since it does not involve anything like the tuning of the regularization parameter $\gamma$. As a matter of fact, the advantage is only apparent because in most cases it is difficult to uniquely specify the prior distribution for the unknown signal and a widespread approach is to estimate from the data at least the variance of the signal. Maximum Likelihood (ML) may be used for this purpose (MacKay 1992, De Nicolao et al. 1997).

The third approach, namely fixed-interval state-space smoothing (Anderson and Moore 1979), can be seen as the specialization of Bayesian regression to problems admitting a state-space representation. The unknown signal is modeled as the output of a finite-dimensional system fed by (continuous-time) white noise and the estimate is worked out by Kalman filtering algorithms in $O(n)$ operations. Also in this case, it is advisable to tune the variance of the so-called process noise using either ML or GCV criteria. For a neural network interpretation of these nonparametric estimators, the interested reader may usefully refer to (Poggio and Girosi 1990, Ferrari-Trecate 1999, De Nicolao and Ferrari-Trecate 2001).

Notably, all the three approaches share the need for the tuning of the amount of smoothing they introduce. This tuning is the nonparametric counterpart of the model order selection problem.
which is encountered when the data are fit by a parametric model (e.g. a polynomial). Most criteria, including ML and GCV, require the repeated evaluation of a positive scalar \( q(\gamma) \) called "equivalent number of parameters" or "equivalent degrees of freedom" (Wahba 1983). The parameter \( q(\gamma) \) ranges from 0 to \( n \) and takes the maximum value when the estimate perfectly interpolates the data. In other words, the equivalent number of parameters measures the flexibility of the nonparametric estimate and provides an indication of how many parameters would be needed by a parametric model to achieve similar approximation properties. By exploiting the analogy with the degrees of freedom of parametric models, the equivalent number of parameters has been used to propose approximate \( \chi^2 \), \( F \)-tests and standard-error bounds (Hastie and Tibshirani 1990).

In Tikhonov regularization and Bayesian regression, the evaluation of \( q(\gamma) \) requires \( O(n^3) \) operations so that the calculation of \( q(\gamma) \) is one of the most computer-intensive part of the overall regression algorithm (Wahba 1990). However, if the problem admits a state-space formulation, \( q(\gamma) \) can be computed in \( O(n) \) operations by averaging the smoothing error covariances obtained through Riccati-type recursions (Kohn and Ansley 1989).

The interest for the equivalent number of parameters goes beyond the issue of numerical efficiency. In particular, the study of the asymptotic properties of \( q(\gamma) \) when \( n \) tends to infinity has been instrumental to establishing statistical consistency and rates of convergence for spline estimates associated with the GCV criterion (Utreras 1983), (Speckman 1985), see also (Silverman 1985). Very recently, spectral factorization concepts have been used to obtain the exact closed-form expression of the asymptotic smoothing ratio \( s(\gamma) \) (defined as the limit of \( q(\gamma)/n \) as \( n \) tends to infinity) of cubic smoothing splines in the uniform sampling case (De Nicolao et al. 2000).

The contribution of the present paper is threefold. First, a new general algorithm is derived for the \( O(n) \) computation of \( q(\gamma) \) via state-space spectral factorization methods. Compared to the most efficient existing method, which required a forward and backward iteration (Kohn and Ansley 1989), the proposed algorithm is faster because it uses only a forward Riccati recursion. When the result is specialized to the uniform sampling case, a fast algorithm for the exact calculation of the asymptotic smoothing ratio \( s(\gamma) \) is obtained whose complexity does not depend on the number of data. Finally, in the cases of linear smoothing splines and first-order deconvolution with first-order regularizer, the closed-form expression of \( s(\gamma) \) is obtained. Beside their obvious application in the development of fast smoothing algorithms, the uniform sampling results have an autonomous interest since they clarify the type of dependence of \( q(\gamma) \) on \( \gamma \) itself and the sampling period \( T \). To make an example, using the results of the paper it is easy to obtain as a function of \( T \) the value of \( \gamma \) ensuring the desired amount of smoothing (measured by \( q(\gamma) \)) without solving linear systems or Riccati equations.

Although the technical machinery is similar, the results of the present paper are much more general than those in (De Nicolao et al. 2000). In fact, herein general state-space smoothing problems are considered instead of restricting the attention to cubic smoothing splines, and the sampling schedule is not restricted to be uniform.
2 Problem statement

Consider the monodimensional regression problem with discrete data given by

\[ y_i = f(t_i) + \epsilon_i, \quad i = 1, \ldots, n \]  \tag{1}

where \( f(\cdot) : \mathbb{R} \rightarrow \mathbb{R} \) is the unknown signal to be reconstructed, \( \epsilon_i \) are zero mean Gaussian random variables independent, identically distributed with variance \( \sigma^2 \) and \( 0 \leq t_1 < t_2 < \ldots < t_n \). Note that, unless specified, the sampling scheduling is in general not uniform. Let \( D_n \) denote the set of the collected datapoints \((t_i, y_i), \quad i = 1, \ldots, n\). In a Bayesian framework, \( f(\cdot) \) is modeled as a stochastic process (the so-called “prior”) that should capture all the a priori information we have about the unknown signal. Then, the aim is to provide an estimate of the unknown signal based on the posterior probability \( P(f|D_n) \). When the prior is a zero-mean Gaussian process (GP), its autocovariance function \( \lambda^2 C(\xi_1, \xi_2) \) (where \( \lambda > 0 \) is a scale parameter) provides a sufficient statistic. Moreover, if \( f(\cdot) \) is independent of the errors \( \epsilon_i \), it follows that \( P(f|D_n) \) is Gaussian with mean and covariance given by (Whittle 1963)

\[
\hat{f}(t) = E[f|D_n](t) = \check{C}'(t)\theta, \tag{2}
\]
\[
\theta = H^{-1}\check{y}, \tag{3}
\]
\[
C_f(\xi_1, \xi_2) = \text{Var}[f|D_n](\xi_1, \xi_2) = \lambda^2 C(\xi_1, \xi_2) - \lambda^2 \check{C}'(\xi_1)H^{-1}\check{C}(\xi_2) \tag{4}
\]
\[
\check{C}(t) = \begin{bmatrix} C(t, t_1) & \cdots & C(t, t_n) \end{bmatrix}', \tag{5}
\]
\[
H = \frac{1}{\lambda^2} \frac{E[\check{y}\check{y}']}{\lambda^2} = \begin{bmatrix} C(t_1, t_1) & \cdots & C(t_1, t_n) \\ \vdots & \ddots & \vdots \\ C(t_n, t_1) & \cdots & C(t_n, t_n) \end{bmatrix} + \frac{\sigma^2}{\lambda^2}I, \quad \check{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \tag{6}
\]

Since the posterior is Gaussian, we may use \( \check{f}(t) \) as a point estimate of the unknown function and obtain error bars from the posterior covariance \( C_f(\xi_1, \xi_2) \). We will refer to \( \check{f}(t) \) as the Bayes estimate. Note that the Bayes estimate is given by the linear combination of the basis functions \( \{C(t, t_i)\}_{i=1}^n \) through the vector of coefficients \( H^{-1}\check{y} \).

In this paper, the signal \( f(\cdot) \) is modeled as the stochastic process generated by the single-input single-output linear time-invariant system

\[
P: \begin{cases} 
\dot{x}(t) = Ax(t) + Bw(t) \\
f(t) = Cx(t) \\
x(0) = 0
\end{cases}, \tag{7}
\]

where \( x(t) \in \mathbb{R}^l \) is the state, \( w(t) \in \mathbb{R} \) is a continuous-time white Gaussian noise with intensity \( \lambda^2 \), and \( A, B, C \) are matrices of suitable dimensions. Since \( P \) is linear and \( w(\cdot) \) is a GP, it
follows that \( f(\cdot) \) is a GP as well (Anderson and Moore 1979). Under the previous assumptions the Bayes estimate may be computed from (2) once the covariance function of \( f(\cdot) \) is derived. The state-space model \( \mathcal{P} \) is very useful in the analysis of mono-dimensional regression problems and it embodies the following notable case.

**Spline problem.** The output of the system \( \mathcal{P} \) with

\[
A = \begin{bmatrix}
0 & 0 & \cdots & 0 \\
1 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 1
\end{bmatrix}, \quad A \in \mathbb{R}^{l \times l}, \quad B = \begin{bmatrix}
1 \\
0 \\
\vdots \\
0
\end{bmatrix}, \quad C = \begin{bmatrix}
0 & \cdots & 0 & 1
\end{bmatrix}
\]

is an \((l - 1)\) fold integrated Wiener process, and, as shown in (Wahba 1990), the Bayes estimate turns out to be a \((2l - 1)\)-th order regression spline with zero boundary conditions.

Consider now deconvolution problems where the data model is

\[
y_k = \int_0^{t_k} h(t_k - \tau)z(\tau)d\tau + \epsilon_k, \quad k = 1, \ldots, n, \quad h(\cdot) \in L^2([0, +\infty])
\]

and the goal is to reconstruct the signal \( z(t) \) from the measurements collected in \( \mathcal{D}_n \). If \( z(\cdot) \) is modeled as a zero-mean GP with autocovariance function \( \lambda^2 Q(\xi_1, \xi_2) \), independent of the errors \( \epsilon_i \), the Bayes estimate is given by the formulas

\[
\dot{z}(t) = \eta(t)\hat{\theta}
\]

\[
\hat{\theta} = \tilde{H}^{-1}\tilde{y}
\]

\[
\eta(t) = \left[ \eta_1(t) \cdots \eta_n(t) \right]
\]

\[
\eta_\tau(t) = \int_0^{+\infty} h(\tau - \xi)Q(t, \xi)d\xi
\]

\[
\tilde{H} = \left[
\begin{array}{ccc}
C(t_1, t_1) & \cdots & C(t_1, t_n) \\
\vdots & \ddots & \vdots \\
C(t_n, t_1) & \cdots & C(t_n, t_n)
\end{array}
\right] + \frac{\sigma^2}{\lambda^2}I
\]

\[
C(\xi_1, \xi_2) = \int_0^{+\infty} \int_0^{+\infty} h(\xi_1 - \tau)h(\xi_2 - t)Q(\tau, t)d\tau dt.
\]

Note that, by defining

\[
f(t) = \int_0^t h(t - \tau)z(\tau)d\tau,
\]

\( f(\cdot) \) is a GP with autocovariance function \( \lambda^2 C(\xi_1, \xi_2) \) and the coefficients \( \hat{\theta} \) are computed in the same way as in (3), i.e. relying only on the statistic of \( f(\cdot) \), the output samples and the noise variance.
In this paper we consider deconvolution problems where $h(t)$ is the impulse response of the SISO linear system

$$\mathcal{F} : \begin{cases} 
\dot{\xi}(t) = \tilde{A}\xi(t) + \tilde{B}z(t) \\
f(t) = \tilde{C}\xi(t) \\
\xi(0) = 0 
\end{cases}, \xi(t) \in \mathbb{R}^r$$

and $z(\cdot)$ is modeled as in the $l$-th order spline case. The signal $f(t)$ is then the output of the system

$$\begin{bmatrix} 
\dot{x}(t) \\
\dot{\xi}(t) 
\end{bmatrix} = \begin{bmatrix} 
A & 0 \\
\tilde{BC} & \tilde{A} 
\end{bmatrix} \begin{bmatrix} 
x(t) \\
\xi(t) 
\end{bmatrix} + \begin{bmatrix} 
B \\
0 
\end{bmatrix} w(t) $$

$$f(t) = \begin{bmatrix} 
0 & \tilde{C} 
\end{bmatrix} \begin{bmatrix} 
x(t) \\
\xi(t) 
\end{bmatrix}$$

that has the same structure of the system (7). Therefore, without loss of generality, the representation (7) will be considered as a state-space realization of the signal $f(\cdot)$ also for deconvolution problems.

A crucial problem in the derivation of the Bayes estimate (both for regression and deconvolution problems) is that, in practice, the value of $\lambda_2$ is seldom known a priori and it (or possibly $\gamma = \sigma^2/\lambda^2$) has to be tuned on the basis of the available data. The most popular criteria to accomplish this goal are the Generalized Cross Validation (GCV) (Wahba 1990) and the Maximum Likelihood (ML) (MacKay 1992). In both of them a central role is played by the so-called influence matrix (Wahba 1990).

**Definition 1.** The uniquely defined $n \times n$ matrix $H(\gamma)$ such that

$$\begin{bmatrix} \hat{f}(t_1) & \hat{f}(t_2) & \ldots & \hat{f}(t_n) \end{bmatrix}' = H(\gamma)\tilde{y}$$

is termed influence matrix.

Defining the sum of squared residuals as

$$SSR(\gamma) = \|(I - H(\gamma))\tilde{y}\|^2,$$

the GCV criteria selects the value of $\gamma$ by minimizing the loss function

$$GCV(\gamma) \doteq \frac{1}{(n - q(\gamma))^2} SSR(\gamma), \quad q(\gamma) \doteq \text{Tr}[H(\gamma)],$$

(for simplicity, the dependence of $q(\gamma)$ and $H(\gamma)$ on the sampling period $T$ and on the number of data is omitted). To accomplish this minimization an iterative strategy is needed involving
the computation of $SSR(\gamma)$ and $q(\gamma)$ for different values of $\gamma$. The bottleneck is the calculation of $q(\gamma)$ which requires $O(n^3)$ operations unless suitable algorithms are employed. The same problem of computing $q(\gamma)$ arises also when the ML criterion is adopted. There exists an extensive literature dealing with the computation of $q(\gamma)$ and by now efficient algorithms are available whose complexity is $O(2n)$, see (Ansley et al. 1993) for a state-space method well as (Wahba 1990, chapter 11) for a review of the various contributions.

It is worth noticing that

$$E[SSR(\gamma)] = (n - q(\gamma))\sigma^2.$$ 

This formula resembles the well known relationship

$$E[SSR] = (n - q)\sigma^2$$

found when trying to fit $n$ noisy data by means of a $q$-th order linear-in-parameters model. By analogy, the real number $q(\gamma)$ has been termed, in (Wahba 1990), equivalent degrees of freedom (EDOF) of the nonparametric estimator $f(t)$. In fact, $q(\gamma)$ ranges from 0 to $n$, as $\gamma$ decreases from $+\infty$ to 0.

The purpose of Section 3 is to derive a new $O(n)$ algorithm for the calculation of $q(\gamma)$ by resorting to spectral factorization techniques. In Sections 4 and 5 we will investigate the behavior of the asymptotic smoothing ratio

$$s(\gamma) \doteq \lim_{n \to +\infty} \frac{q(\gamma)}{n},$$

when the sampling schedule is uniform. In particular we will derive an algorithm for the evaluation of $s(\gamma)$ whose complexity is independent of the number of data points and, in some relevant special cases, we will provide the closed-form expression of $s(\gamma)$.

### 3 Spectral characterization of the equivalent degrees of freedom

The rationale of our derivation is as follows. First it is shown that $q(\gamma)$ depends only on the statistics of the sampled process $f(t_i)$. Then, we find a state-space representation for such a process. Moreover Kalman filtering is used to show that $q(\gamma)$ only depends on the variance of the output of a suitable linear time-varying system $W^*$. Finally, this variance is computed via an $O(n)$ matrix recursive equation.

**Lemma 2.** Let

$$\bar{f}_d \doteq \begin{bmatrix} f(t_1) & f(t_2) & \cdots & f(t_n) \end{bmatrix}^\prime$$

and $\Gamma \doteq E[\bar{f}_d \bar{f}_d^\prime]/\lambda_2$. Then,

$$H(\gamma) = (I + \gamma \Gamma^{-1})^{-1}.$$ 

**Proof.** From (2)-(6) it is apparent that it holds that
\[
\begin{bmatrix}
\hat{f}(t_1) & \cdots & \hat{f}(t_n)
\end{bmatrix} = \mathbb{E}[\tilde{f}_a \tilde{f}_d] \mathbb{E}[yy']^{-1} \tilde{y} = \lambda^2 \Gamma \left( \lambda^2 \Gamma + \sigma^2 I \right)^{-1} \tilde{y} = (I + \gamma \Gamma^{-1})^{-1} \tilde{y},
\]

and the result follows from the definition of influence matrix.

**Definition 3.** Let \( L_S \in \mathbb{R}^{n \times n} \) be a lower triangular matrix and consider the linear time-varying system

\[
\begin{cases}
x_{k+1} = A_k x_k + B_k u_k \\
y_k = C_k x_k + D_k u_k, x_k \in \mathbb{R}^l.
\end{cases}
\]

Then, the 4-uple \( S = (A_k, B_k, C_k, D_k) \) is said a realization of \( L_S \) if,

\[
\begin{bmatrix}
y_1 & y_2 & \cdots & y_n
\end{bmatrix} = L_S \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}, \tag{12}
\]

for all the sequences \( u_k \). Conversely, given the quadruple \( S = (A_k, B_k, C_k, D_k) \), the associated matrix \( L_S \) such that (12) holds is termed \((n\text{-step})\) input-output matrix of the system.

**Definition 4.** Let \( L_S \in \mathbb{R}^{n \times n} \) be a lower triangular matrix and consider the linear time-varying system

\[
\begin{cases}
x_{k+1} = A_k x_k + B_k u_k \\
y_k = C_k x_k , x_k \in \mathbb{R}^l.
\end{cases}
\]

Then, the triple \( S = (A_k, B_k, C_k) \) is said a (strictly proper) realization of \( L_S \) if,

\[
\begin{bmatrix}
y_1 & y_2 & \cdots & y_n
\end{bmatrix} = L_S \begin{bmatrix} u_0 & u_1 & \cdots & u_{n-1} \end{bmatrix}, \tag{13}
\]

for all the sequences \( u_k \). Conversely, given the triple \( S = (A_k, B_k, C_k) \), the associated matrix \( L_S \) such that (13) holds is termed \((n\text{-step strictly proper})\) input-output matrix of the system.

Note that the input-output matrices are uniquely defined and they do not depend on the input sequence. If the associated system \( S \) is time-invariant, the matrix \( L_S \) takes a Toeplitz structure.

In the sequel, the notation \( w_k \sim WGN(\lambda^2) \) means that the sequence \( w_k \) is a discrete-time Gaussian white noise with intensity \( \lambda^2 \).

**Lemma 5.** Let \( L_W \in \mathbb{R}^{n \times n} \) be the \((n\text{-step strictly proper})\) input-output matrix of the system \( W = (F_k, G_k, C) \) where \( F_k = e^{A(t_{k+1}-t_k)} \) and \( G_k \) is such that

\[
G_k G_k' = \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-t)} B B' e^{A'(t_{k+1}-t)} dt. \tag{14}
\]

Then \( \Gamma = L_W L_W' \).
Proof. See (De Nicolao et al. 2000), Lemma 2.

As corollary of Lemma 5, it turns out that the linear time-varying system

\[
\begin{aligned}
    x_{k+1} &= F_k x_k + G_k w_k \\
    \hat{y}_k &= C x_k + e_k \\
    x_0 &= 0
\end{aligned}
\]

where \( w(k) \in \mathbb{R}^{n \times 1} \), \( w_k \sim WGN(\lambda^2 I) \), is a stochastic realization of the discrete-time process \( y_k \) defined by (1), (7), i.e. \( \hat{y}_k \) has the same statistics of \( y_k \).

**Lemma 6.** Let \((\Gamma + \gamma I)^{-\frac{1}{2}}\) be a lower triangular matrix s.t.

\[
(\Gamma + \gamma I)^{-1} = \left[ (\Gamma + \gamma I)^{-\frac{1}{2}} \right]' (\Gamma + \gamma I)^{-\frac{1}{2}}
\]

and \( L_W \) be defined as in Lemma 5. Then it holds

\[
q(\gamma) = \text{Tr} \left[ (\Gamma + \gamma I)^{-\frac{1}{2}} L_W L_W' \left[ (\Gamma + \gamma I)^{-\frac{1}{2}} \right]' \right]
\]

Proof. See (De Nicolao et al. 2000), Lemma 3.

The next step is to find a suitable system that is a realization of \((\Gamma + \gamma I)^{-\frac{1}{2}}\). The answer is provided by the next Lemma.

**Lemma 7.** The system

\[
K = (F_k - K_k C, K_k, -C(R_k^c)^{-\frac{1}{2}}, (R_k^c)^{-\frac{1}{2}}),
\]

where

\[
\begin{aligned}
    R_k^c &= C \Sigma_k C' + \sigma^2 I, \\
    K_k &= F_k \Sigma_k C' \left[ C \Sigma_k C' + \sigma^2 \right]^{-1}, \\
    \Sigma_{k+1} &= F_k \Sigma_k F_k' - F_k \Sigma_k C' \left[ C \Sigma_k C' + \sigma^2 \right]^{-1} C \Sigma_k F_k' + \lambda^2 G_k G_k', \\
    \Sigma_0 &= 0,
\end{aligned}
\]

is a realization of the matrix \((\Gamma + \gamma I)^{-\frac{1}{2}}\).

Proof. See (De Nicolao et al. 2000), Lemma 4.

At this point, it is apparent that a realization of the input-output matrix \((\Gamma + \gamma I)^{-\frac{1}{2}} L_W \) is the cascade \( W^* \) of the systems \( W \) and \( K \), as depicted in Fig. 1.
Fig. 1. The system $\mathcal{W}^*$

Note that $\mathcal{W}^* = (F^*_k, G^*_k, C^*_k)$ where

$$
F^*_k = \begin{bmatrix}
F_k & 0 \\
K_k C & F_k - K_k C
\end{bmatrix},
G^*_k = \begin{bmatrix}
G_k \\
0
\end{bmatrix},
C^*_k = (R^*_k)^{-\frac{1}{2}} \begin{bmatrix}
C & -C
\end{bmatrix}.
$$

If this system is fed with $\tilde{w}_k \sim WGN(1)$ and

$$
\begin{bmatrix}
\eta_1 \\
\eta_2 \\
\vdots \\
\eta_n
\end{bmatrix}' = \left(\Gamma + \gamma I\right)^{-\frac{1}{2}} L_W \left[\tilde{w}_1 \ \tilde{w}_2 \ \cdots \ \tilde{w}_n\right]',
$$

from Lemma 6 follows that,

$$
q(\gamma) = \sum_{k=1}^{n} \text{Var}[\eta_k]. \quad (16)
$$

In other words $\mathcal{W}^*$ is a spectral factor for the signal $\eta_k$ that is involved in the calculation of (16). Rather remarkably, the computation of (16) may be carried out in a simple way as shown in the next theorem.

**Theorem 8.** The equivalent degrees of freedom are given by

$$
q(\gamma) = \sum_{k=1}^{n} C P_k C',
$$

where the matrices $R^*_k$, $F_k$ are as in (15), $F_k = e^{A(t_{k+1} - t_k)}$ and $P_k$ satisfies

$$
P_{k+1} = (F_k - K_k C)P_k(F_k - K_k C)' + \lambda^2 (R^*_k)^2 G_k G_k',
\quad P_0 = 0. \quad (17)
$$

**Proof.** We apply to $\mathcal{W}^*$ the state-space change of coordinates (that does not alter the output $\eta_k$)

$$
T = \frac{1}{2} \begin{bmatrix}
I & -I \\
I & I
\end{bmatrix}, \quad T^{-1} = \begin{bmatrix}
I & I \\
-I & I
\end{bmatrix}
$$

so obtaining a new system $(\tilde{F}_k, \tilde{G}_k, \tilde{C}_k)$ defined by
\[
\begin{align*}
\tilde{F}_k &= T F_k^* T^{-1} = \begin{bmatrix} F_k - K_k C & 0 \\ K_k C & F_k \end{bmatrix}, \\
\tilde{G}_k &= T^{-1} G_k^* = \frac{1}{2} \begin{bmatrix} G_k \\ G_k \end{bmatrix} \\
\tilde{C}_k &= C_k^* T = (R_k^*)^{-\frac{1}{2}} \begin{bmatrix} 2C & 0 \end{bmatrix}
\end{align*}
\]

In this new representation it is apparent that an equivalent realization of \((\tilde{F}_k, \tilde{G}_k, \tilde{C}_k)\) is the system

\[
\begin{align*}
\tilde{x}_{k+1} &= (F_k - K_k C)\tilde{x}_k + (R_k^*)^{-\frac{1}{2}}\tilde{w}_k \\
\eta_k &= C\tilde{x}_k \\
\tilde{x}_0 &= 0
\end{align*}
\]

Then the thesis follows from the fact that, as shown (Anderson and Moore 1979) for a general linear system, the variance of the output \(\eta_k\) may be computed as \(\text{Var}[\eta_k] = CP_k C'\) where \(P_k\) satisfies (17).

The algorithm for the calculation of the equivalent degrees of freedom is summarized in the following steps:

(1) The computation of \(F_k\) and \(G_k\) as defined in the system \(\mathcal{W}\).
(2) The \(n\) updates of the matrices \(R_k^*, K_k, \text{and } \Sigma_k\) as in (15) and of the matrix \(P_k\) as in (17)
(3) The computation of \(q(\gamma)\) via (16).

The second step is the computational bottleneck of the algorithm and it is worth noticing that it takes \(O(n)\) operations. The main advantage of our algorithm with respect to the one proposed in (Ansley et al. 1993) is that the former requires only a forward Kalman filtering whereas the latter needs also a backward Kalman filtering that doubles the computational cost.

4 The uniform sampling case: Analysis of the asymptotic smoothing ratio

If the sampling is uniform, the data model becomes

\[
y_k = f(kT) + e_k, \ T > 0, \ k = 1, ..., n.
\]

The system \(\mathcal{W}\) of Lemma 5 turns out to be time invariant since

\[
\begin{align*}
F_k &= F = e^{AT} \\
G_k G' &= GG' = \int_0^T e^{A(T-t)} BB' e^{A'(T-t)} dt.
\end{align*}
\]
However the Kalman filter

\[ K_k = (F - K_kC, K_k, -C(R_k^e)^{-\frac{1}{2}}, (R_k^e)^{-\frac{1}{2}}), \]

\[ R_k^e = C\Sigma_k C' + \sigma^2 I \]

\[ K_k = F\Sigma_k C'[C\Sigma_k C' + \sigma^2]^{-1} \]

\[ \Sigma_{k+1} = \Sigma_k - F\Sigma_k C' [C\Sigma_k C' + \sigma^2]^{-1}C\Sigma_k F' + \lambda^2 GG' \]

\[ \Sigma_0 = 0 \]

is always time-varying.

Our next aim is the analysis of the asymptotic smoothing ratio that, in view of (16) may be computed as

\[ s(\gamma) = \lim_{n \to +\infty} \frac{1}{n} \sum_{k=1}^{n} \text{Var}[\eta_k] \]  

(19)

where \( \eta_k \) is the output of the system \( W^* \) fed with \( \tilde{w}_k \sim WGN(1) \).

Since we are interested in the asymptotic of \( \eta_k \), the behavior of \( K \) for \( k \to +\infty \) will be first analyzed.

**Assumption A** The continuous-time pair \( (A, B) \) is reachable and the discrete-time pair \( (F, C) \) is observable.

**Lemma 9.** Let Assumption A hold. Then the following facts hold:

1. For all possible nonnegative definite initial conditions \( \Sigma_0, \lim_{k \to +\infty} \Sigma_k = \Sigma \) and \( \lim_{k \to +\infty} K_k = K \), where \( K = F\Sigma C'[C\Sigma C' + \lambda]^{-1} \), \( \Sigma \) being the unique nonnegative definite symmetric solution of the algebraic Riccati equation \( \Sigma = \Sigma - F\Sigma C'[C\Sigma C' + \lambda]^{-1}C\Sigma F' + GG' \);

2. All the poles of the transfer function matrix \( W_K(z)W(z) \) have modulus strictly less than one, where

\[ W_K(z) = [I - C(zI - F + KC)^{-1}K] (C\Sigma C' + \lambda I)^{-\frac{1}{2}} \]

\[ W(z) = C(zI - F)^{-1}G. \]

**Proof.** See (De Nicolao et al. 2000), Lemma 5. □

We are now in a position to give an integral formulation of the asymptotic smoothing ratio.

**Theorem 10.** Let \( \Phi(z) = W(z)W(z^{-1}) \). Then, under the Assumption A,

\[ s(\gamma) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi(e^{j\omega})}{\Phi(e^{j\omega}) + \gamma} d\omega. \]  

(20)

**Proof.** See (De Nicolao et al. 2000), Theorem 1. □
The integral (20) plays a key role in the fast algorithm for the evaluation of the equivalent degrees of freedom we propose. This procedure may be summarized as follows.

First step: calculation of $GG'$. The integral in (18) can be easily evaluated applying the method proposed in (Van Loan 1978) that leads $GG' = \Phi_1'\Psi$, with

$$F = \begin{bmatrix} -A & b' \\ 0 & A \end{bmatrix}, \quad e^F = \begin{bmatrix} \Phi_1 & \Psi \\ 0 & \Phi_2 \end{bmatrix}, \quad \Phi_1 \in \mathbb{R}^{l \times l}.$$  

Second step: spectral factorization. First of all, find the matrix $G$ according the Cholesky decomposition of $GG'$. Then, evaluate $\Phi(z) = \hat{G}(z)\hat{G}(z^{-1})$, where $\hat{G}(z) = C(zI - e^{AT})^{-1}G$. Last, compute the polynomials $A(z)$ and $B(z)$ that satisfy

$$\frac{\Phi(z)}{\Phi(z) + \gamma} = \frac{B(z)B(z^{-1})}{A(z)A(z^{-1})}. \quad (21)$$

Third step: Åström algorithm. The integral in (20) can be rewritten as

$$\frac{1}{2\pi i} \oint_{|z|=1} \frac{B(z)B(z^{-1})}{A(z)A(z^{-1})} \, dz, \quad (22)$$

and can be evaluated applying the Åström algorithm (Åström 1970) that uses only of the $A(z)$’s and $B(z)$’s coefficients. This algorithm requires also the stability of the $A(z)$ polynomial (i.e. its roots belong to the open unit ball). In view of the following Lemma, this assumption is always met with.

Lemma 11. Let $\Phi(z) = N_\Phi(z)/D_\Phi(z)$, where $N_\Phi(z)$ and $D_\Phi(z)$ are coprime polynomials. Then, $\forall z : |z| = 1$, the denominator of (21) is non null.

Proof. It is easy to verify that $N_\Phi(e^{j\omega}) \geq 0$, $D_\Phi(e^{j\omega}) \geq 0$, $\omega \in [0, 2\pi)$ By contradiction, if the denominator of (21) is null for some $\tilde{\omega} \in [0, 2\pi)$, this implies $N_\Phi(e^{j\tilde{\omega}}) = -\gamma D_\Phi(e^{j\tilde{\omega}})$. If both sides are null, then $N_\Phi(z)$ and $D_\Phi(z)$ possess a common root and they are not coprime, otherwise the positivity of $N_\Phi(e^{j\tilde{\omega}})$ does not hold. 

Note that the computational cost of the above procedure is independent of the datapoints number.

5 Explicit formulas for deconvolution and smoothing splines

In this Section, the procedure for the computation of the asymptotic smoothing ratio is applied to the deconvolution and smoothing spline problems. In the basic cases, closed-form formulas for $s(\gamma, T)$ are derived.
Specializing the algorithm of Section 4 to the first order spline problem \((l = 1)\), one obtains \(GG' = T\) and, by setting \(\xi = \gamma/T\),

\[
\Phi(z) = \frac{\lambda^2 T}{(z - 1)(z^{-1} - 1)}, \quad B(z) = 1,
\]

\[
A(z) = a_0 z + a_1, \quad a_0 = -\frac{\xi}{a_1}, \quad a_1 = \frac{1}{2} - \sqrt{\frac{1}{4} + \xi}.
\]

Then, the final formula is

\[
s(\gamma, T) = s(\xi) = \frac{\sqrt{1 + 4\xi} - 1 - 2\xi}{1 + 4\xi - \sqrt{1 + 4\xi} (1 + 2\xi)}.
\]

A log-scale plot of the function \(s(\xi)\) is given in Figure 2. This result is notable under various aspects. First of all it shows that \(s(\gamma, T)\) depends on the ratio \(\gamma/T\). Hence, when considering experiments with different sampling rates, the same value of the smoothing ratio could correspond to extremely different values of \(\gamma\), so that \(\gamma\) alone is not a meaningful index of smoothness. To make an example, one cannot assume that \(\gamma = 10^7\) corresponds to a smoothing ratio close to zero if the value of \(T\) is ignored. In this respect Figure 2 is very helpful because it provides a link between \(\gamma\) and \(q(\gamma)\). Vice-versa, it is also possible to quickly obtain the value of \(\gamma\) that guarantees the desired smoothing ratio.

Now we consider the case of the deconvolution problem with

\[
A = \begin{bmatrix} 0 & 0 \\ \rho & a \end{bmatrix}, \quad c = \begin{bmatrix} 0 & 1 \end{bmatrix}
\]
where $a > 0$. Applying the algorithm, we derive

$$\Phi(z) = \frac{\rho^2 T^3(z + 4 + z^{-1})}{6(z - 1)(z^{-1} - 1)(z - e^{-aT})(z^{-1} - e^{-aT})}$$

and, by setting $\hat{\xi} = \frac{\gamma}{\rho^2 T^3}$, the asymptotic smoothing ratio turns out to be

$$\hat{b}_1 = \pm \sqrt{\frac{3}{2}} \pm \sqrt{\frac{1}{2}}$$

$$\hat{a}_1 = \frac{\sqrt{6} - \sqrt{2 + 24\xi(e^{-aT} + 1)^2}}{2}$$

$$\hat{a}_0 = \frac{1 - 6\xi(e^{-aT + 1})^2}{a_1} + \frac{\sqrt{\left(1 - 6\xi(e^{-aT + 1})^2\right)^2 - 24\xi e^{-aT}}}{2}$$

$$\hat{s}(\gamma, T) = \hat{s}(\hat{\xi}) =$$

$$\frac{\hat{a}_0^2}{\hat{b}_1^2 \left(\hat{a}_0 - (6\xi e^{-aT})^2\right)^2} \left[\hat{b}_1^4 + \frac{(\hat{a}_0^2 + 6\xi e^{-aT} - \hat{b}_1^2 \hat{a}_0 \hat{a}_1)^2}{(\hat{a}_0^2 + 6\xi e^{-aT})^2 - \hat{a}_0^2 \hat{a}_1^2}\right]$$

Moreover, when $\rho = 1$, by setting $\hat{\xi} = \frac{\gamma}{T^3}$, we can obtain the asymptotic smoothing ratio $\tilde{s}(\hat{\xi})$ for the third-order smoothing spline problem as $\tilde{s}(\hat{\xi}) = \lim_{a \to 0} \hat{s}(\hat{\xi})$. The closed-form formula of $\tilde{s}(\hat{\xi})$ is reported in (De Nicolao et al. 2000). The log-plots of $\tilde{s}(\xi)$ for various values of $a$ and $\tilde{s}(\xi)$, are shown in Figure 3.
Acknowledgments

This paper has been partially supported by MURST Project “Identification and Control of Industrial Systems”, and ICE-CNR.

References


